Available online at www.sciencedirect.com

ScienceDirect

International Journal of Approximate Reasoning
46 (2007) 346–365

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

www.elsevier.com/locate/ijar

Complexity measurement of fundamental pseudo-independent models

J. Lee *, Y. Xiang

University of Guelph, Department of Computing and Information Science, Guelph, Ont., Canada N1G 4S7

Received 1 July 2005; received in revised form 1 April 2006; accepted 21 September 2006
Available online 16 November 2006

Abstract

Pseudo-independent (*PI*) *models* are a special class of *probabilistic domain model* (*PDM*) where a set of marginally independent domain variables shows *collective dependency*, a special type of dependency associated with the scope of a set of variables in a probabilistic domain. Due to this property, common probabilistic learning methods based on a *single-link lookahead search* cannot learn PI models. To learn PI models, a learning algorithm should be equipped with a search with its scope beyond a single link, which is called a *multi-link lookahead search*. An improved result can be obtained by incorporating model complexity into a scoring metric to explicitly trade off model accuracy for complexity and vice versa during selection of the best model among candidates at each learning step. To implement this scoring metric for learning PI models, the complexity formula for every class of PI models is required. Previous studies found the complexity formula for *full PI models*, one of the three major types of PI models (the other two are *partial* and *mixed PI models*). This study presents the complexity formula for *atomic partial PI models*, partial PI models that contain no embedded PI submodels. This paper shows the complexity can be acquired by arithmetic operation with the cardinality of the space of domain variables in an atomic partial PI model. The new formula provides the basis for further characterizing the complexity of *non-atomic PI models*, which contain embedded PI submodels in their domains.

Keywords: Probabilistic reasoning; Knowledge discovery; Data mining; Machine learning; Belief networks; Model complexity

* Corresponding author.
E-mail addresses: jaelee11@hotmail.com (J. Lee), yxiang@cis.uoguelph.ca (Y. Xiang).

0888-613X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2006.09.008

1. Introduction

Learning probabilistic networks [1–4] has been an active area of research recently. The task of learning such networks is NP-hard [5]. Therefore, learning algorithms use a heuristic search, and a common search heuristic is a single-link lookahead which generates network structures that differ by a single link at each level of the search.

Pseudo-independent (*PI*) models [6] are a class of probabilistic domain models where a group of marginally independent variables displays collective dependency. PI models can be classified into three types: *full*, *partial*, and *mixed* PI models based on the pattern of marginal independency and the range of collective dependency. The most restrictive type is full PI model where every proper subset of variables is marginally independent. This restriction is relaxed in partial PI models where every proper subset of variables is not necessarily marginally independent. While, in full or partial PI models, all variables in the domain are collectively dependent, this is not the case in mixed PI models where only proper subsets, called *embedded PI subdomains*, of variables are collectively dependent. In mixed PI models, however, the marginal independency pattern of each embedded PI subdomain is that of either a full or a partial PI model.

PI models cannot be learned by a single-link lookahead search, which cannot recover underlying collective dependency. Incorrectly learned models introduce silent errors when used for decision making. To learn PI models, a more sophisticated search method called *multi-link lookahead* [7] should be used. It was implemented in a learning algorithm called *RML* [8]. The algorithm is equipped with the Kullback–Leibler cross entropy as the scoring metric for the score of the goodness-of-fit to data, or model accuracy.

The scoring metric of the learning algorithm can be improved by combining the cross entropy with a measure of model complexity. A simple way of combination is a weighted sum of the two measures, but other alternatives such as using logarithm functions are also possible. By using such a scoring metric, between two models of the same accuracy that can be measured by cross entropy, the one with lesser complexity gives a higher score and will be preferred. The focus of this paper is on the assessment of model complexity.

Model complexity is defined by the number of parameters required to fully specify a probabilistic domain model. In previous work [9], a formula was presented for estimating the number of parameters in *full PI models*, one of the three types of PI models (the other two are *partial* and *mixed PI models*). However, the formula was very complex, and did not show the structural dependence relationships among parameters. The new concise formula for full PI models was recently presented [10] using a *JPD hypercube* [11], a hyper-dimensional extension of a JPD table for visualizing the dependency relations among parameters.

In this study, we present the model complexity formula for *atomic partial PI models*, a class of partial PI model that contains no embedded PI submodels. Atomic partial PI models are the building blocks of mixed PI models. This new formula is simple in form, and provides insight into the structural dependency relationship among parameters of partial PI models. Furthermore, the previous complexity formula for full PI models is integrated into this new formula. In other words, substituted with the conditions of full PI models, the new formula is reduced to the previous formula for full PI models; this confirms that full PI models are a special case of partial PI model whose partition blocks are all singleton. In addition, the new formula provides the basis for further characterizing the complexity of mixed PI models. We apply the hypercube method to show how the complexity of partial PI models is acquired from the spaces of variables.

2. Background

Let V be a set of n discrete variables $X_1, \ldots, X_n$—in what follows we will focus on domains of finite and discrete variables. Each variable X_i has a finite space $S_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,D_i}\}$ of cardinality D_i. The space of a set V of variables is defined by the Cartesian product of the space of every variable in V, that is, $S_V = S_1 \times \cdots \times S_n$ (or $\prod_i S_i$). Thus, S_V consists of tuples of all possible combinations in values of variables in V. Each tuple is called a *configuration* of V, denoted by $(x_1, \ldots, x_n)$.

Let $P(X_i)$ denote the probability function over X_i and $P(x_i)$ denote the probability value $P(X_i = x_i)$. The following axiom of probability [12] is called the *total probability law*:

Definition 1 (*Total probability law*). For any variables X_i with its finite space $S_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,D_i}\}$ of cardinality D_i,

$$P(S_i) = P(x_{i,1}) + P(x_{i,2}) + \cdots + P(x_{i,D_i}) = 1. \tag{1}$$

One of the properties of a probability function is *marginalization*:

Definition 2 (*Marginalization*). For $S_i = \{x_{i,1}, x_{i,2}, \ldots, x_{i,D_i}\}$ and any subsets A of V,

$$P(x_{i,1}, A) + P(x_{i,2}, A) + \cdots + P(x_{i,D_i}, A) = P(A). \tag{2}$$

This process is called marginalization. The set of variables $V \backslash A$, where $\backslash$ is the operator for set difference, that is eliminated after summation, such as $\{X\}$ in (2), is said to be marginalized out variables.

Note that all the above axioms and properties apply not only to a single variable X_i, but to any subsets of V, which can be thought as a single *composite variable* made up of a subset of elementary variables. In this is the case, the space of the subset or the composite variable is defined as the *joint space* of all *elementary variables* in the subset, that is, the Cartesian product of the space of every elementary variable.

Definition 3 (*Probabilistic domain model*). A probabilistic domain model (PDM) $\mathscr{M}$ over V defines the probability value of every configuration for every subset $A \subseteq V$.

Definition 4 (*Joint probability*). For a set of discrete random variables $V = \{X_1, \ldots, X_n\}$, let $P(V)$ or $P(X_1, \ldots, X_n)$ denote the joint probability distribution (JPD) function of $\{X_1, \ldots, X_n\}$. We refer to $P(x_1, \ldots, x_n)$, or the joint probability value of a configuration $(x_1, \ldots, x_n)$, as a joint parameter, or simply a "joint". A table made up of entries of all joint parameters for explicitly specifying a JPD is called a JPD table or joint table.

Therefore, a JPD model, represented by a JPD table over a domain, is a PDM.

There are more than a single method to specify a PDM. An intuitive method is direct specification, where a probability value, i.e., a parameter, is assigned to every configuration of the states of all domain variables.

With the JPD of a domain, the parameter for any configurations over a subset of variables can be derived by marginalization (Eq. (2)), which sums up all joint parameters that contain the configuration wanted to be obtained. For example, consider a domain

$V = \{X_1, X_2, X_3\}$ of three ternary variables. A parameter $P(x_{2,1}, x_{3,2})$, for instance, can be obtained by summing up all joint parameters that contain the configuration $(x_{2,1}, x_{3,2})$:

$$P(x_{2,1}, x_{3,2}) = P(x_{1,1}, x_{2,1}, x_{3,2}) + P(x_{1,2}, x_{2,1}, x_{3,2}) + P(x_{1,3}, x_{2,1}, x_{3,2}).$$

Definition 5 (*Marginal probability*). We refer to the probability function $P(A)$ over a proper subset $A \subset V$ as marginal probability distribution (MPD) of A, and its probability value as a marginal parameter or simply a "marginal" since it is obtained by marginalization.

Definition 6 (*Dependent, and independent parameters*). In direct specification of a PDM, a probability value that can be derived from other probability values by using constraint rules on the domain, such as axioms of probability, is called a dependent, or constrained parameter. A probability value that cannot be derived in such a manner is called an independent or free or unconstrained parameter.

Definition 7 (*Model complexity*). The number of independent parameters required to fully specify a PDM is referred to as the model complexity of the PDM, denoted as ω.

The following lemma on the number of independent parameters for specifying the probability distribution of a variable (or a composite variable) is easily derived from the total probability law in Definition 1.

Lemma 8 (*Number of independent parameters in MPD*). *Given a variable X_i with the cardinality of its space D_i, the number of independent parameters for fully specifying the marginal probability distribution $P(X_i)$ is upper-bounded by*

$$D_i - 1. \tag{3}$$

When no information of the constraints on a general PDM is given, the PDM should be specified only by joint parameters. The required number of joint parameters is given as follows.

Lemma 9 (*Number of independent parameters for a JPD model*). *Given a domain of n variables with the cardinality of the space of each variable $D_1, \ldots, D_n$, the number of independent parameters for fully specifying the JPD model on the domain is upperbounded by*

$$\omega_g = \left(\prod_{i=1}^{n} D_i\right) - 1. \tag{4}$$

Proof. The total number of joint parameters in direct specification is $\prod_{i=1}^{n} D_i$. However, the number of parameters required is one less because, with $n - 1$ parameters, one parameter can be derived (it is, thus, dependent) by the total probability law in Definition 1. □

Definition 10 (*Conditional independency*). For any three disjoint subsets A, B, and C of V, the sets A and B are said to be conditionally independent given C, denoted by $I(A, B|C)$, if the following equality holds for all values of the variables in A, B, and C:

$$P(A|B, C) = P(A|C), \quad \textit{where } P(B, C) > 0. \tag{5}$$

Definition 11 (*Marginal independency*). Any two disjoint subsets A and B in V are said to be marginally independent, written $I(A, B|\varnothing)$, if the following equality holds for all values of the variables in A and B:

$$P(A|B) = P(A) \quad \textit{where } P(B) > 0. \tag{6}$$

Remark. Marginal independency is a special case of conditional independency where $C = \varnothing$ in (5).

Corollary 12 (JPD of marginally independent variable). *For marginally independent variables* $X_1, \ldots, X_n$,

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i). \tag{7}$$

Definition 13 (*Generally dependent*). Variables in a set A are said to be generally dependent if, for every proper subset $B \subset A$,

$$P(B|A \setminus B) \neq P(B). \tag{8}$$

Definition 14 (*Collectively dependent*). Variables in a set A are *said to be* collectively dependent if, for each proper subset $B \subset A$, there exists no proper subset $C \subset A \setminus B$ that satisfies

$$P(B|A \setminus B) = P(B|C). \tag{9}$$

Collective dependency prevents conditional independency and modeling through any proper subsets of variables.

Table 1 shows an example of collectively dependent variables. In this example, A consists of four binary variables $A = \{X_1, X_2, X_3, X_4\}$. Their marginals are

$$P(x_{1,0}) = 0.48, \quad P(x_{2,0}) = 0.56, \quad P(x_{3,0}) = 0.66, \quad P(x_{4,0}) = 0.54.$$

Table 1
Collectively dependent variables

(X_1, X_2, X_3, X_4)	$P(X_1, X_2, X_3, X_4)$	(X_1, X_2, X_3, X_4)	$P(X_1, X_2, X_3, X_4)$
(0,0,0,0)	0.12	(1,0,0,0)	0.09
(0,0,0,1)	0.14	(1,0,0,1)	0.07
(0,0,1,0)	0.03	(1,0,1,0)	0.05
(0,0,1,1)	0.01	(1,0,1,1)	0.05
(0,1,0,0)	0.06	(1,1,0,0)	0.11
(0,1,0,1)	0.03	(1,1,0,1)	0.04
(0,1,1,0)	0.04	(1,1,1,0)	0.04
(0,1,1,1)	0.05	(1,1,1,1)	0.07

Consider $B = \{X_1\}$ for a proper subset $B \subset A$ for instance. Then $A \backslash B = \{X_2, X_3, X_4\}$. There exists no proper subset $C \subset A \backslash B$ that satisfies $P(B|A \backslash B) = P(B|C)$, as shown in what follows for a case of $\{x_{1,0}, x_{2,0}, x_{3,0}, x_{4,0}\}$:

$$P(x_{1,0}|x_{2,0}, x_{3,0}, x_{4,0}) = 0.57,$$

which is not the same as

$$P(x_{1,0}|x_{2,0}, x_{3,0}) = 0.63, \quad P(x_{1,0}|x_{2,0}, x_{4,0}) = 0.52, \quad P(x_{1,0}|x_{3,0}, x_{4,0}) = 0.47$$

or

$$P(x_{1,0}|x_{2,0}) = 0.54, \quad P(x_{1,0}|x_{3,0}) = 0.46, \quad P(x_{1,0}|x_{4,0}) = 0.56.$$

Note that, in this example, variables in A are also generally dependent.

The following definition shows some properties of the conditional independency relation found in [13–16].

Definition 15 (*Symmetry and composition of conditional independency*). Let A, B, C, D be subsets of V. The following properties of conditional independency are called symmetry and composition, respectively:

- Symmetry:

$$I(A, B|C) \iff I(B, A|C). \tag{10}$$

- Composition:

$$I(A, B|C) \wedge I(A, D|C) \Rightarrow I(A, B \cup D|C). \tag{11}$$

Composition is recognized by Pearl [14] (Eq. (3.34b)) as a necessary property of a DAG-isomorph, and is known as a *typical* property of graphical models [17]. Fig. 1, depicting Table 2, shows a domain with the composition property. The nodes represent the variables in the table, and the edges express dependency. In this domain, both $I(H, M|C)$ and $I(H, F|C)$ hold, and, hence, $I(H, M \cup F|C)$ holds.

However, the composition property does not hold for every graphical model. For instance, PI models do not follow this property [18]. An example is given in Fig. 2 where a complete graph represents collective dependency. For example, $I(X_3, X_2 \cup X_4|X_1)$ does not hold while both $I(X_3, X_2|X_1)$ and $I(X_3, X_4|X_1)$ hold, because collective dependency exists among all four variables $X_1, \ldots, X_4$.

A *pseudo-independent* (*PI*) model is a PDM where proper subsets of a set of collectively dependent variables display marginal independency [7].

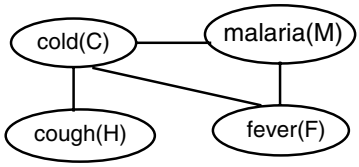

Fig. 1. The undirected graphical model on the JPD of Table 2.

Table 2
The JPD table for the domain of Fig. 1

(C,H,F,M)	$P(C,H,F,M)$	(C,H,F,M)	$P(C,H,F,M)$
(0,0,0,0)	0.785813	(1,0,0,0)	0.012737
(0,0,0,1)	0.000041	(1,0,0,1)	0.000002
(0,0,1,0)	0.041359	(1,0,1,0)	0.002248
(0,0,1,1)	0.000787	(1,0,1,1)	0.000014
(0,1,0,0)	0.068332	(1,1,0,0)	0.007218
(0,1,0,1)	0.000004	(1,1,0,1)	0.000009
(0,1,1,0)	0.003596	(1,1,1,0)	0.012737
(0,1,1,1)	0.000068	(1,1,1,1)	0.000077

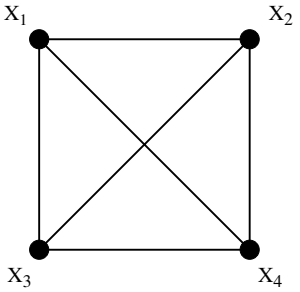

Fig. 2. The undirected graphical model on the JPD of Table 1.

Definition 16 (*Full PI model*). A PDM over a set V ($|V| \geqslant 3$) of variables is a full PI model if the following properties (called axioms of full PI models) hold:

(S_I) Variables in every proper subset of V are marginally independent.
(S_{II}) Variables in V are collectively dependent.

Table 1 is also an example of such PI domains. With collective dependency of the domain, which is already shown above, any proper subsets of A are marginally independent, e.g.,

$$P(x_{1,1}, x_{2,1}, x_{3,0}) = P(x_{1,1})P(x_{2,1})P(x_{3,0}) = 0.15.$$

The complexity of full PI models is given as follows:

Theorem 17 (Complexity of full PI models [10]). *Let a PDM $\mathscr{M}$ be a full PI model over $V = \{X_1, \ldots, X_n\}$. Then the number of independent parameters of M is upper-bounded by*

$$\omega_f = \prod_{i=1}^{n}(D_i - 1) + \sum_{i=1}^{n}(D_i - 1). \tag{12}$$

The axiom (S_I) of marginal independency is relaxed in partial PI models, as is defined through *marginally independent partition*.

Definition 18 (*Marginally independent partition*). Let $V(|V| \geqslant 3)$ be a set of variables and $B = \{B_1, \ldots, B_m\}$ $(m \geqslant 2)$ be a partition of V. B is a marginally independent partition if for every subset $A = \{X_{i,k} | X_{i,k} \in B_k \text{ for } k = 1, \ldots, m\}$, variables in A are marginally independent. Each partition block B_k in B is called a marginally independent block.

A marginally independent partition of V groups variables in V into m marginally independent blocks. The property of marginally independent blocks is that if a subset A is formed by taking one element from different blocks, then variables in A are always marginally independent.

In a *partial PI model*, it is not necessary that every proper subset is marginally independent as is the case of a full PI model.

Definition 19 (*Partial PI model*). A PDM over a set $V(|V| \geqslant 3)$ of variables is a partial PI model on V if the following properties (called axioms of partial PI models) holds:

(S_I') V can be partitioned into two or more marginally independent blocks.
(S_{II}) Variables in V are collectively dependent.

The following definitions on *maximum marginally independent partition* is needed later for obtaining the complexity of partial PI models:

Definition 20 (*Maximum partition and minimum block*). Let $B = \{B_1, \ldots, B_m\}$ be a marginally independent partition of a partial PI model over V. B is a maximum marginally independent partition if there exists no marginally independent partition B' of V such that $|B| < |B'|$. The blocks of a maximum marginally independent partition are called the minimum marginally independent blocks or minimum blocks.

For visualizing the analysis of dependency relationship among parameters in a PI model, a graphical representation of parameters called a *JPD hypercube* is introduced. A JPD hypercube is a hyper-dimensional extension of a JPD table. This perspective is useful when a systematic counting of joint parameters is necessary.

Definition 21 (*JPD hypercube*). A structural arrangement of the joint parameters in a JPD table into a hyper-dimensional space is called a JPD hypercube or simply a hypercube. The dimension of a hypercube is defined as the number of distinct variables in a PDM.

Consider a general PDM $\mathscr{M}$ over a set of n variables $V = \{X_1, \ldots, X_n\}$. The JPD table of $\mathscr{M}$ contains $\prod_{i=1}^{n} D_i$ entries. This number equals the number of parameters in the JPD.

Given a PDM $\mathscr{M}$ over $\{X_1, \ldots, X_n\}$, the corresponding hypercube has n-dimensions with axes $X_1, \ldots, X_n$. An axis X_i consists of *segments* of unit length and that are labeled by each value of X_i, e.g., $x_{i,1}, x_{i,2}, \ldots, x_{i,D_i}$. Therefore, the length of each axis X_i is D_i. A *cell* is a basic element of a hypercube, which has exactly $\prod_{i=1}^{n} D_i$ cells, each of which corresponds to a joint parameter (or an entry in a JPD table). The cell located at $X_1 = x_{i,j}, X_2 = x_{2,k}, \ldots, X_n = x_{n,m}$ is labeled by the joint parameter $P(X_1 = x_{i,j}, X_2 = x_{2,k}, \ldots, X_n = x_{n,m})$ or simply $p(j, k, \ldots, m)$.

As an example, Fig. 5a shows a hypercube for a PDM with three variables X_1, X_2, and X_3, shown in Fig. 4. Variable X_1 and X_2 are ternary, and X_3 is binary. The hypercube for this PDM contains 18 cells, each of which represents a joint parameter. The cell labeled by

$p(2,3,1)$ represents the joint parameter for the probability $P(X_1 = x_{1,2}, X_2 = x_{2,3}, X_3 = x_{3,1})$ and is located in the hypercube at the point where the three planes $X_1 = x_{1,2}$, $X_2 = x_{2,3}$, and $X_3 = x_{3,1}$ meet, as pointed in Fig. 5a.

3. Complexity of atomic partial PI models

The following defines *atomic* partial PI models:

Definition 22 (*Atomic PI model*). A PI model $\mathscr{M}$ over a set V ($|V| \geqslant 3$) of variables is an atomic PI model if no collective dependency exists in any proper subsets of V.

Table 3 shows the JPD of an atomic PI model depicted in Fig. 3. In the figure, each minimal marginally independent block is represented by a dotted circle such as $\{X_1, X_2\}$, $\{X_3\}$, $\{X_4, X_5\}$. Solid lines inside a circle depicts the marginal dependency among the variables in a block. For example, there is a solid link between X_1 and X_2 since X_1 and X_2 are marginally dependent. On the other hand, collective dependency other than marginal dependency is depicted by dashed lines.

As shown in the table and figure, the domain V consists of five binary variables $\{X_1, X_2, X_3, X_4, X_5\}$, and the marginals are

$$P(x_{1,0}) = 0.6585, \quad P(x_{2,0}) = 0.4055, \quad P(x_{3,0}) = 0.3988,$$
$$P(x_{4,0}) = 0.4450, \quad P(x_{5,0}) = 0.6663.$$

The maximum marginally independent partition consists of three minimal blocks $\{\{X_1, X_2\}, \{X_3\}, \{X_4, X_5\}\}$. For example, X_2 is marginally independent of each variable in the other blocks, e.g.,

$$P(x_{2,0}, x_{3,1}) = P(x_{2,0})P(x_{3,1}) = 0.1617$$

and

$$P(x_{2,1}, x_{5,0}) = P(x_{2,1})P(x_{5,0}) = 0.3961.$$

Table 3
An atomic partial PI model

$(X_1, X_2, X_3, X_4, X_5)$	$P(X_1, \ldots, X_5)$	$(X_1, X_2, X_3, X_4, X_5)$	$P(X_1, \ldots, X_5)$
(0,0,0,0,0)	0.03694	(1,0,0,0,0)	0.02957
(0,0,0,0,1)	0.02171	(1,0,0,0,1)	0.02029
(0,0,0,1,0)	0.05086	(1,0,0,1,0)	0.04507
(0,0,0,1,1)	0.02228	(1,0,0,1,1)	0.01710
(0,0,1,0,0)	0.02286	(1,0,1,0,0)	0.02127
(0,0,1,0,1)	0.01605	(1,0,1,0,1)	0.01181
(0,0,1,1,0)	0.03539	(1,0,1,1,0)	0.02825
(0,0,1,1,1)	0.01313	(1,0,1,1,1)	0.01300
(0,1,0,0,0)	0.07105	(1,1,0,0,0)	0.02644
(0,1,0,0,1)	0.04647	(1,1,0,0,1)	0.01507
(0,1,0,1,0)	0.10489	(1,1,0,1,0)	0.03571
(0,1,0,1,1)	0.04166	(1,1,0,1,1)	0.01606
(0,1,1,0,0)	0.04879	(1,1,1,0,0)	0.01589
(0,1,1,0,1)	0.02918	(1,1,1,0,1)	0.01165
(0,1,1,1,0)	0.06794	(1,1,1,1,0)	0.02534
(0,1,1,1,1)	0.02929	(1,1,1,1,1)	0.00900

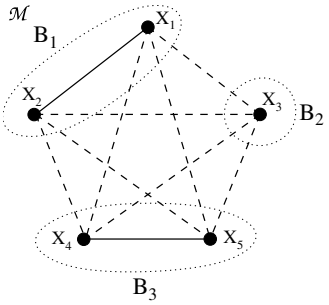

Fig. 3. A non-atomic partial PI model of the JPD shown in Table 3.

On the other hand, the variables in the same blocks are marginally dependent, e.g., $P(x_{1,1}, x_{2,1}) = 0.1552$ which is different from $P(x_{1,1})P(x_{2,1}) = 0.2030$, and $P(x_{4,0}, x_{5,1}) = 0.1722$ which is different from $P(x_{4,0})P(x_{5,1}) = 0.1485$. All five variables are collectively dependent: For example,

$$P(x_{3,0}|x_{1,1}x_{2,0}, x_{4,0}, x_{5,1}) = 0.6321;$$

however,

$$P(x_{3,0}|x_{2,0}, x_{4,0}, x_{5,1}) = 0.6012, \quad P(x_{3,0}|x_{1,1}, x_{4,0}, x_{5,1}) = 0.6012,$$
$$P(x_{3,0}|x_{1,1}, x_{2,0}, x_{5,1}) = 0.6012, \quad P(x_{3,0}|x_{1,1}, x_{2,0}, x_{4,0}) = 0.6012.$$

Similarly,

$$P(x_{1,1}|x_{2,0}, x_{3,1}, x_{4,1}, x_{5,1}) = 0.4974;$$

however,

$$P(x_{1,1}|x_{3,1}, x_{4,1}, x_{5,1}) = 0.3415, \quad P(x_{1,1}|x_{2,0}, x_{4,1}, x_{5,1}) = 0.4595,$$
$$P(x_{1,1}|x_{2,0}, x_{3,1}, x_{5,1}) = 0.4595, \quad P(x_{1,1}|x_{2,0}, x_{3,1}, x_{4,1}) = 0.4595.$$

Because of the dependency within each block, the complexity of partial PI models is higher than full PI models, but lower than general PDMs with variables of the same space cardinalities, as we will see later.

The following lemma states that in a PDM that satisfies composition (Eq. (11)), if every pair of variables between two subsets are marginally independent, then the two subsets are marginally independent.

Lemma 23 (*Marginal independency of subsets*). *Let $\mathcal{M}$ be a PDM over V where composition holds in every subset. Let $B_\alpha = \{Y_1, \ldots, Y_s\}$ and $B_\beta = \{Z_1, \ldots, Z_t\}$ denote any two disjoint non-empty subsets of variables in V. If $I(Y_i, Z_j|\varnothing)$ holds for every pair (Y_i, Z_j), then*

$$I(B_\alpha, B_\beta|\varnothing). \tag{13}$$

Proof. We prove that $I(Y_i, Z_j|\varnothing)$ for every (Y_i, Z_j) implies $I(Y_i, B_\beta|\varnothing)$ and that $I(Y_i, B_\beta|\varnothing)$ for every Y_i implies $I(B_\alpha, B_\beta|\varnothing)$.

Applying composition recursively from $I(Y_i, Z_1|\varnothing)$ to $I(Y_i, Z_t|\varnothing)$ gives $I(Y_i, Z_1 \cup \cdots \cup Z_t|\varnothing)$ or $I(Y_i, B_\beta|\varnothing)$. By symmetry (Eq. (10)), $I(Y_i, B_\beta|\varnothing)$ is equivalent to $I(B_\beta, Y_i|\varnothing)$. In the

same manner, by applying composition recursively from $I(B_\beta, Y_1|\emptyset)$ to $I(B_\beta, Y_s|\emptyset)$ gives $I(B_\beta, Y_1 \cup \cdots \cup Y_s)$ or $I(B_\beta, B_\alpha|\emptyset)$. By symmetry this is equivalent to $I(B_\alpha, B_\beta|\emptyset)$. □

Lemma 23 assumes that the PDM satisfies composition. Composition implies no collective dependency in the PDM as follows:

Lemma 24 (Composition implies no collective dependency). *Let $\mathscr{M}$ be a PDM over V that satisfies composition. Then no collective dependency exists in V.*

Proof. This lemma directly follows from the definition of composition (Eq. (11)). □

The following Lemmas 25 and 26 are required for proving Lemma 27 on marginal independency of blocks. The two lemmas state that the marginal independency of partition is preserved after removing a proper subset of variables from the partition blocks and after merging the blocks. □

Lemma 25 (*Marginal independency of any subpartitions*). *Let a PDM $\mathscr{M}$ be a partial PI model over $V = \{X_1, \ldots, X_n\}$ $(n \geqslant 3)$ with a marginally independent partition $B = \{B_1, \ldots, B_m\}$ $(m \geqslant 2)$. Consider a new partition $B' = \{B'_1, \ldots, B'_m\}$ over V' $(V' \subseteq V)$ which is defined by removing a proper subset of variables from one or more partition blocks such that $B'_i \subseteq B_i$ for every B'_i $(i = 1, \ldots, m)$. Then, B' is also a marginally independent partition called a subpartition.*

Proof. By the above assumption, $B'_k \subseteq B_k$ $(k = 1, \ldots, m)$. Consider every subset $A' = \{Y_{i,k}|Y_{i,k} \in B'_k \text{ for } k = 1, \ldots, m\}$. Then for each A' it is clear that $Y_{i,k} \in B_k$ holds for $k = 1, \ldots, m$. Therefore, by the definition of marginally independent partition (Definition 18), B' is also a marginally independent partition. □

Lemma 26 (Marginal independency after merging). *Let a PDM $\mathscr{M}$ be a partial PI model over $V = \{X_1, \ldots, X_n\}$ $(n \geqslant 3)$ with a marginally independent partition $B = \{B_1, \ldots, B_m\}$ $(m \geqslant 2)$. Consider a new partition $\widehat{B} = \{\widehat{B}_1, \ldots, \widehat{B}_r\}$ $(r < m)$ over V which is defined by merging one or more blocks in B. Then, $\widehat{B}$ is also a marginally independent partition.*

Proof. Consider every subset $A = \{X_{i,k}|X_{i,k} \in B_k \text{ for } k = 1, \ldots, m\}$. For every A, a set $\widehat{A}$ can be always found such that $\widehat{A} = \{Y_{i,q}|Y_{i,q} \in \widehat{B}_q \text{ for } q = 1, \ldots, r\}$ and $A \subseteq \widehat{A}$. The reason is, for every B_k $(k = 1, \ldots, m)$, a block $\widehat{B}_j$ in $\widehat{B}$ can be always found that satisfies $B_k \subseteq \widehat{B}_j$. From the definition of marginally independent partition (Definition 18), every variable in A is marginally independent. Therefore, every variable in $\widehat{A}$ is also marginally independent and, thus, $\widehat{B}$ is also a marginally independent partition. □

The following lemma states that in an atomic partial PI model, removing one variable from V makes any two blocks in V marginally independent.

Lemma 27 (Marginal independency between two independent blocks). *Let a PDM $\mathscr{M}$ be an atomic partial PI model over $V = \{X_1, \ldots, X_n\}$ $(n \geqslant 3)$, where composition holds in every proper subset. Let a marginally independent partition be denoted by $B = \{B_1, \ldots, B_m\}$. When $m > 2$, for any two distinct blocks B_r and B_q in B,*

$$I(B_r, B_q|\varnothing). \tag{14a}$$

When $m = 2$, *that is* $B = \{B_1, B_2\}$, *for any* $X_i \in B_1$ *or any* $X_j \in B_2$,

$$I(B_1 \setminus \{X_i\}, B_2|\varnothing) \quad \text{and} \quad I(B_1, B_2 \setminus \{X_j\}|\varnothing). \tag{14b}$$

Proof. Statement (14a) and (14b) hold due to Lemma 23 and the fact that in a proper subset of an atomic PI domain, composition holds and collective dependency does not. □

Consider the domain in Table 3. The maximum marginally independent partition consists of three minimal blocks $B_1 = \{X_1, X_2\}$, $B_2 = \{X_3\}$, and $B_3 = \{X_4, X_5\}$. The number $m = 3$ which is greater than 2, and therefore by Eq. (14a), $I(B_1, B_2|\varnothing)$, $I(B_1, B_3|\varnothing)$, and $I(B_2, B_3|\varnothing)$ hold. This is demonstrated as follows: First, from $I(B_1, B_2|\varnothing)$, there should be $P(B_1, B_2) = P(B_1)P(B_2)$. From Table 3, we have $P(x_{1,0}, x_{2,1}, x_{3,0}) = P(x_{1,0}, x_{2,1})P(x_{3,0}) = 0.2641$. Next, from $I(B_1, B_3|\varnothing)$, there should be $P(B_1, B_3) = P(B_1)P(B_3)$. Numerically, we have $P(x_{1,0}, x_{2,1}, x_{4,1}, x_{5,0}) = P(x_{1,0}, x_{2,1})P(x_{4,1}, x_{5,0}) = 0.1728$. Finally, from $I(B_2, B_3|\varnothing)$, we should have $P(B_2, B_3) = P(B_2)P(B_3)$, which is verified by $P(x_{3,1}, x_{4,1}, x_{5,0}) = P(x_{3,1})P(x_{4,1}, x_{5,0}) = 0.1569$.

As shown in the parameterization of full PI models [10], a sum of joint probabilities can be represented by a product of marginal probabilities. While the size of the joint space grows exponentially on the number of variables, the total size of marginal spaces grows only linearly. This results in the lesser complexity of a full PI model compared with a general PDM. A similar joint–marginal relationship holds for partial PI models. However, the marginal probability is from each partition block, not from each variable as is the case with full PI models. This is shown in the following lemma:

Lemma 28 (*Joint–marginal equality of atomic PI models*). *Let a PDM* $\mathscr{M}$ *be an atomic partial PI model over* $V = \{X_1, \ldots, X_n\}$ $(n \geqslant 3)$, *where composition holds in every proper subset. Let a marginally independent partition be denoted by* $B = \{B_1, \ldots, B_m\}$ $(m \geqslant 2)$. *For an arbitrary* X_i, *let* $B' = \{B'_1, \ldots, B'_m\}$ *denote a subpartition of* B *made by removing an* X_i *from* B *so that every* B'_i *is the same as* B_i *except the block from which* X_i *was removed. Then*,

$$\sum_{k=1}^{D_i} P(X_1, \ldots, x_{i,k}, \ldots, X_n) = P(B'_1) \cdots P(B'_m). \tag{15}$$

Proof. The summation $\sum_{k=1}^{D_i} P(X_1, \ldots, x_{i,k}, \ldots, X_n)$ at the left represents marginalization on a variable X_i. This is equal to $P(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ or in partitional notation $P(B'_1, \ldots, B'_m)$.

The proof by induction is as follows:

- Base case where $m = 2$: We need to show $P(B'_1, B'_2) = P(B'_1)P(B'_2)$. This is equivalent to $I(B'_1, B'_2|\varnothing)$, which is obvious by Lemma 27.
- Induction hypothesis: Assume the following holds for $m = k$:

$$P(B'_1, \ldots, B'_k) = P(B'_1) \cdots P(B'_k). \tag{16}$$

- Induction step: We need to show the following holds for $m = k + 1$:

$$P(B'_1, \ldots, B'_{k+1}) = P(B'_1) \cdots P(B'_{k+1}).$$

Merging $\{B'_1, \ldots, B'_k\}$ into one block and applying Lemmas 25–27 give

$$P(B'_1, \ldots, B'_k, B'_{k+1}) = P(B'_1, \ldots, B'_k)P(B'_{k+1}).$$

By the induction hypothesis (Eq. 16),

$$P(B'_1, \ldots, B'_k)P(B'_{k+1}) = P(B'_1) \cdots P(B'_k)P(B'_{k+1}).$$

Therefore, by mathematical induction from $m = 2, k$ and $k + 1$, Eq. (15) must hold for any m. □

Lemma 28 can be verified with the domain in Table 3 whose maximum marginally independent partition consists of three minimal blocks $B_1 = \{X_1, X_2\}$, $B_2 = \{X_3\}$, and $B_3 = \{X_4, X_5\}$. Suppose X_5 is removed from the domain. Then by Eq. (15), there should be

$$P(X_1, X_2, X_3, X_4) = P(B'_1)P(B'_2)P(B'_3) = P(X_1, X_2)P(X_3)P(X_4),$$

which is verified by

$$P(x_{1,0}, x_{2,1}, x_{3,0}, x_{4,1}) = P(x_{1,0}, x_{2,1})P(x_{3,0})P(x_{4,1}) = 0.1466.$$

Corollary 29, which directly follows from Lemma 28, shows the relation between joint parameters and marginal parameters, by which joint parameters can be derived from other marginal and joint parameters.

Corollary 29 (Atomic PI joint constraint). *Let a PDM $\mathscr{M}$ be an atomic partial PI model over $V = \{X_1, \ldots, X_n\}$ $(n \geqslant 3)$, where composition holds in every proper subset. Let a marginally independent partition be denoted by $B = \{B_1, \ldots, B_m\}$ $(m \geqslant 2)$. For an arbitrary X_i, let $B' = \{B'_1, \ldots, B'_m\}$ denote a subpartition of B made by removing an X_i from B so that every B'_i is the same as B_i except the block from which X_i was removed. Then,*

$$P(X_1, \ldots, x_{i,r}, \ldots, X_n) = P(B'_1) \cdots P(B'_m) - \sum_{k=1, k \neq r}^{D_i} P(X_1, \ldots, x_{i,k}, \ldots, X_n). \tag{17}$$

We are to derive the number of independent parameters for specifying a PDM by using the constraint (Corollary 29) on atomic partial PI domains. First, we determine the number of independent marginal parameters, denoted as ω_m, that are required for specifying all $P(B'_1), \ldots, P(B'_m)$ terms in Corollary 29. Next, we determine the number of joint parameters that cannot be derived from the ω_m marginal parameters and other joint parameters. Using a hypercube representation, this amounts to counting the number of cells that cannot be derived from others since a cell in a hypercube corresponds to a joint parameter in the JPD. The procedure is as follows:

(1) Check cells one by one to see whether it can be derived from the ω_m marginal parameters and other cells by applying Corollary 29.
(2) Since we are determining the number of underivable cells, as soon as a cell is determined to be derivable, it is eliminated from the hypercube and from further consideration.

Repeat this procedure until no more cells can be eliminated further. The remaining cells and the ω_m marginal parameters constitute independent parameters of the partial PI model.

For example, consider the partial PI model in Fig. 4, which corresponds to the hypercube in Fig. 5a. The PDM consists of three variables $\{X_1, X_2, X_3\}$; X_1, X_2 are ternary, and

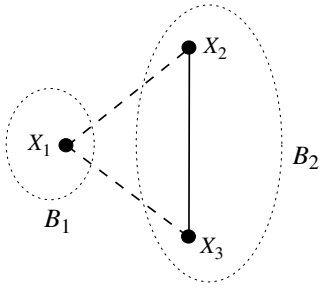

Fig. 4. A partial PI model. (The dotted circle depicts each partition block.)

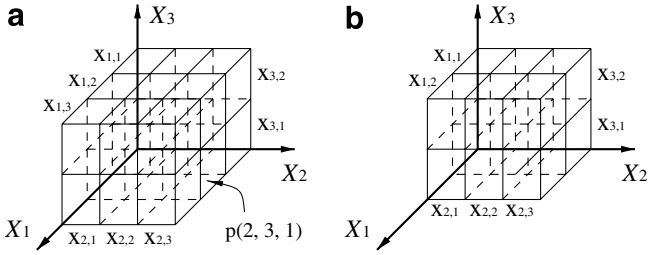

Fig. 5. Eliminating derivable cells from a JPD hypercube. (a) The original hypercube ($3 \times 3 \times 2$). (b) The hypercube after eliminating $X_1 = x_{1,3}$.

X_3 is binary. The marginally independent partition $B = \{B_1, B_2\}$ or $\{\{X_1\}, \{X_2, X_3\}\}$. For this PDM, $\omega_m = (3-1) + (3 \times 2 - 1) = 7$. For example, to specify $P(X_1)$ for B_1, two marginals are required such as

$$P(x_{1,1}), P(x_{1,2});$$

and to specify $P(X_2, X_3)$ for B_2, five parameters are required such as

$$P(x_{2,1}, x_{3,1}), P(x_{2,2}, x_{3,1}), P(x_{2,3}, x_{3,1}), P(x_{2,1}, x_{3,2}), P(x_{2,2}, x_{3,2}). \tag{18}$$

We assume that the seven marginal parameters have been specified, and thus the other two marginal parameters $P(x_{1,3})$ and $P(x_{2,3}, x_{3,2})$ can be derived by the total probability law (Eq. (1)). Then, by marginalization, the following six marginal parameters can also be derived from the five parameters in (18) plus $P(x_{2,3}, x_{3,2})$:

$$P(x_{2,1}), P(x_{2,2}), P(x_{2,3}), P(x_{3,1}), P(x_{3,2}), P(x_{3,3}).$$

We refer to the set of cells with the identical value $X_i = x_{i,j}$ as the *hyperplane* at $X_i = x_{i,j}$ in the hypercube. For example, the hyperplane at $X_1 = x_{1,3}$ in Fig. 5a refer to the following six cells, where we use $p(i,j,k)$ as an abbreviation for $P(X_{1,i}, X_{2,j}, X_{3,k})$:

$$p(3,1,1), p(3,2,1), p(3,3,1), p(3,1,2), p(3,2,2), p(3,3,2).$$

By Corollary 29, we have

$$p(3,1,1) = P(x_{2,1}, x_{3,1}) - (p(1,1,1) + p(2,1,1)).$$

That is, the cell at the front-lower-left corner can be derived by the two cells behind it and the marginal parameters. All other cells on the hyperplane at $X_1 = x_{1,3}$ can be similarly derived. Therefore, these six cells can be eliminated from further consideration. The remaining 12 cells are shown in Fig. 5b.

Using the same idea, four of the remaining cells at $X_2 = x_{2,3}$ can be derived, and therefore eliminated from further consideration. The remaining eight cells are shown in Fig. 6a. Again, four of the remaining cells at $X_3 = x_{3,2}$ can be derived. After eliminating them, only four cells are left, as shown in Fig. 6b:

$$p(1,1,1), p(2,1,2), p(1,2,1), p(2,2,1).$$

Since no more cells can be further eliminated, the number of independent parameters needed to specify the partial PI model is 11, with seven marginal parameters and four joint parameters. Note that it would take 17 parameters to specify the JPD of a general PDM over three variables of the same space cardinalities.

The following lemma says that the maximum partition is unique.

Lemma 30 (Unique maximum partition). *Every partial PI model has a unique maximum marginally independent partition.*

Proof. Suppose the maximum marginally independent partition for a partial PI model $\mathscr{M}$ is not unique. Let B and B' denote the two different maximum partitions. Then there must be two variables X_i and X_j that are in the same block B_k in B but in different blocks in B'. Because X_i and X_j are in different partition blocks in B', $I(X_i, X_j|\emptyset)$ must hold. This means the partition block B_k in B can be further partitioned into at least two marginally independent partitions: one that contains X_i, and the other that contains X_j. This contradicts the assumption that B and B' are maximum marginally independent partition. Therefore, Lemma 30 must be true. □

The following is necessary for proving Lemma 34.

Lemma 31 (Variables in minimal blocks are marginally dependent). *Let a partial PI model $\mathscr{M}$ over a set V have the maximum marginally independent partition $B = \{B_1, \ldots, B_m\}$. Then the variables in any minimal blocks $B_1, \ldots, B_m$ are marginally dependent themselves.*

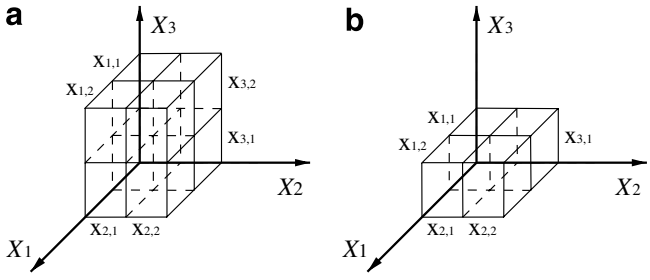

Fig. 6. Eliminating derivable cells from a JPD hypercube. (a) The hypercube after eliminating $X_2 = x_{2,3}$. (b) The hypercube after eliminating $X_3 = x_{3,2}$.

Proof. We prove by contradiction. Suppose a pair of variables X and Y in B_i $(i = 1, \ldots, m)$ are marginally independent. Then B_i can be split into two marginally independent blocks: one containing X and the other containing Y. This contradicts the original assumption that B_i is a minimal block. Therefore, Lemma 31 is true. □

The following states that elements in a minimal block cannot be split into two or more distinct marginally independent blocks of a partition:

Lemma 32 (Variables from distinct minimal blocks cannot form a marginally independent block). *Let a partial PI model $\mathscr{M}$ over a set V have the maximum marginally independent partition $B = \{B_1, \ldots, B_m\}$. Let C be a subset of V such that there exists $B_\alpha \in B$ where $B_\alpha \cap C \neq \emptyset$ and $B_\alpha \not\subset C$. Then C cannot be a block of any marginally independent partition.*

Proof. We prove by contradiction. Suppose that C is a block $\widehat{B}_1$ of a marginally independent partition $\widehat{B} = \{\widehat{B}_1, \ldots, \widehat{B}_q\}$ over V. Then by the definition of marginally independent partition (Definition 18), for every subset $A = \{X_{i,k} \mid X_{i,k} \in \widehat{B}_k \text{ for } k = 1, \ldots, q\}$, variables in A should be marginally independent. Consider A which includes the following two marginally independent elements: one element X from $\widehat{B}_1 \cap B_\alpha$ and another element Y from $B_\alpha \setminus \widehat{B}_1$, since $X \in \widehat{B}_1$ and $Y \in \widehat{B}_j$ where $\widehat{B}_j \in \{\widehat{B}_2, \ldots, \widehat{B}_q\}$. However, since $X, Y \in B_\alpha$, the two elements X and Y are marginally dependent due to Lemma 31, which is a contradiction. □

The following lemma that directly follows from Lemma 32 is needed to prove Lemma 34.

Lemma 33 (Every non-minimal block consists of minimal blocks). *Let a partial PI model $\mathscr{M}$ over a set V have the maximum marginally independent partition $B = \{B_1, \ldots, B_m\}$ $(m \geqslant 2)$. Let $\widehat{B} = \{\widehat{B}_1, \ldots, \widehat{B}_r\}$ $(r < m)$ denote a partition that is not the maximum partition. Then at least one block of $\widehat{B}$ consists two or more minimal blocks.*

The following states that the number of independent marginal parameters required for specifying the joint space of each block of a partition is the minimum when the partition is the maximum partition:

Lemma 34 (The smallest number of independent marginals if partition is the maximum). *Let a PDM $\mathscr{M}$ be a partial PI model over $V = \{X_1, \ldots, X_n\}$ $(n \geqslant 3)$ with a marginally independent partition $B = \{B_1, \ldots, B_m\}$ $(m \geqslant 2)$. Let ω_m denote the total number of marginal parameters required for specifying the joint space of each block $B_1, \ldots, B_m$ of B. Then ω_m is the minimum when B is the maximum partition.*

Proof. Let $B = \{B_1, \ldots, B_m\}$ $(m \geqslant 2)$ denote the maximum marginally independent partition. Let $\widehat{B} = \{\widehat{B}_1, \ldots, \widehat{B}_r\}$ $(r < m)$ denote a partition that is not the maximum partition. Then due to Lemma 33, at least one block $\widehat{B}_i$ consists of two or more minimal blocks. Let the minimal blocks that $\widehat{B}_i$ includes be notated with the block index in the superscript of $\widehat{B}_i$. For example if $\widehat{B}_i$ include minimal blocks B_2, B_4, B_6, then it is notated as $\widehat{B}_i^{(2,4,6)}$. Let k denote the number of minimal blocks that $\widehat{B}_i$ includes. Then the proof by induction from $k = 2$ to $k = m$ is as follows:

- Base case where $k = 2$: Let the two minimal blocks be denoted by B_1 and B_2, and the cardinality of the joint space of each block B_1, B_2, and $\widehat{B}_i^{(1,2)}$ be denoted by D_{B_1}, D_{B_2}, and $D_{\widehat{B}_i^{(1,2)}}$, respectively. The number of marginal parameters for specifying the joint space of each block B_1, B_2, and $\widehat{B}_i^{(1,2)}$ can be represented by $D_{B_1} - 1$, $D_{B_2} - 1$, and $D_{\widehat{B}_i^{(1,2)}} - 1$, respectively. $(D_{\widehat{B}_i^{(1,2)}} - 1)$ is always greater than $(D_{B_1} - 1) + (D_{B_2} - 1)$ because
$$(D_{\widehat{B}_i^{(1,2)}} - 1) - [(D_{B_1} - 1) + (D_{B_2} - 1)] = (D_{B_1} \cdot D_{B_2} - 1) - [(D_{B_1} - 1) + (D_{B_2} - 1)]$$
$$= (D_{B_1} - 1)(D_{B_2} - 1),$$
which is greater than zero since $(D_{B_1} - 1) > 0$ and $(D_{B_2} - 1) > 0$.
- Induction hypothesis: Assume the following holds for $k = q$.
$$(D_{\widehat{B}_i^{(1,\ldots,q)}} - 1) - [(D_{B_1} - 1) + \cdots + (D_{B_q} - 1)] > 0.$$
- Induction step: We need to show the following holds for $k = q + 1$:
$$(D_{\widehat{B}_i^{(1,\ldots,q+1)}} - 1) - [(D_{B_1} - 1) + \cdots + (D_{B_{q+1}} - 1)] > 0.$$

Since $D_{\widehat{B}_i^{(1,\ldots,q+1)}} = D_{\widehat{B}_i^{(1,\ldots,q)}} \cdot D_{B_{q+1}}$, the right-hand side of the above can be written as

$$(D_{\widehat{B}_i^{(1,\ldots,q)}} \cdot D_{B_{q+1}} - 1) - [(D_{B_1} - 1) + \cdots + (D_{B_{q+1}} - 1)]$$
$$= D_{B_{q+1}}(D_{\widehat{B}_i^{(1,\ldots,q)}} - 1) - [(D_{B_1} - 1) + \cdots + (D_{B_q} - 1)].$$

Since $D_{B_{q+1}} > 1$ and from the induction hypothesis, this is greater than zero.

This implies that if a partition contains at least one block that is other than a minimal block (in other words, the partition is not the maximum partition), then the number of independent marginal parameters required for specifying the joint space of each block of the partition is greater than the case of the maximum partition. □

Now we present the general result on the number of independent parameters of atomic partial PI models.

Theorem 35 (Complexity of atomic partial PI models). *Let a PDM $\mathscr{M}$ be an atomic partial PI model over $V = \{X_1, \ldots, X_n\}$ ($n \geqslant 3$), where composition holds in every proper subset. Let $D_1, \ldots, D_n$ denote the cardinality of the space of each variable. Let the maximum marginally independent partition of V be denoted by $B = \{B_1, \ldots, B_m\}$ ($m \geqslant 2$), and the cardinality of the space of each minimal block $B_1, \ldots, B_m$ be denoted by $D_{(B_1)}, \ldots, D_{(B_m)}$, respectively. Then, the number ω_{ap} of parameters required for specifying the JPD of $\mathscr{M}$ is upper-bounded by*

$$\omega_{ap} = \prod_{i=1}^{n}(D_i - 1) + \sum_{j=1}^{m}(D_{(B_j)} - 1). \tag{19}$$

Proof. Before proving the theorem, we explain the result briefly. The first term on the right is the cardinality of the joint space of a general PDM over the set of variables except the space of each variable is reduced by one. This term is the same as the one in the model complexity formula for the full PI model (Eq. (17)). The second term is the number of marginal parameters for specifying the joint space of each minimal partition block. This number is the smallest number required for specifying the marginals of blocks as shown in Lemma 34.

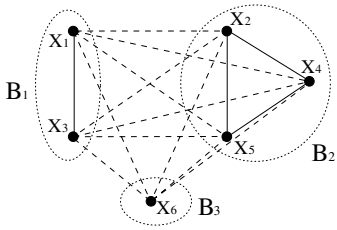

Fig. 7. A partial PI model with three blocks and a total of six variables. (The dotted circles depict each partition block.)

What we need to show is how to derive all joint parameters with $\sum_{j=1}^{m}(D_{(B_j)}-1)$ marginals plus $\prod_{i=1}^{n}(D_i-1)$ joints. First, $\sum_{j=1}^{m}(D_{(B_j)}-1)$ marginal parameters are required for specifying all $P(B'_1),\ldots,P(B'_m)$ terms in Corollary 29. We construct a JPD hypercube for $\mathscr{M}$ to apply Eq. (17) among groups of cells. Applying Corollary 29 and using the similar argument for the example in Fig. 4, we can eliminate hyperplanes at $X_1 = x_{1,D_1}, X_2 = x_{2,D_2},\ldots,X_n = x_{n,D_n}$ in that order such that for each X_i, all cells on the hyperplanes at $X_i = x_{i,D_i}$ can be derived from cells outside the hyperplane and the marginal parameters. The remaining cells form a hypercube whose length along the X_i axis is $D_i - 1$ $(i = 1,2,\ldots,n)$. Therefore, the total number of cells in this hypercube is $\prod_{i=1}^{n}(D_i-1)$. □

The following shows how to use Theorem 35 to compute complexity of an atomic partial PI model.

Example 36 (*Computing model complexity*). Consider the partial PI model in Fig. 7. The domain consists of six variables from X_1 to X_6, and X_1, X_2, X_3 are ternary; X_4, X_5 are binary; and X_6 is 5-nary. The domain has a marginally independent partition $\{B_1, B_2, B_3\}$ or $\{\{X_1, X_3\}, \{X_2, X_4, X_5\}, \{X_6\}\}$.

The number of marginal parameters for all partition blocks is 23, given by $(3 \cdot 3 - 1) + (3 \cdot 2 \cdot 2 - 1) + (5 - 1)$. The number of independent joint parameters is 32, given by $(2-1)^2(3-1)^3(5-1)$. Therefore, the total number of parameters for specifying the domain in this example is $23 + 32 = 55$. Compare this number with the number of parameters for specifying a general PDM over the same set of variables by using the total probability law, giving $2^2 \cdot 3^3 \cdot 5 - 1 = 539$. This shows that the complexity of a partial PI model is significantly less than that of a general PDM.

Note that Theorem 35 holds also for full PI models since a full PI model is a special case of partial PI models. The proof of this is done by substituting $D_{(B_j)}$ in Eq. (19) with D_i (every partition block of full PI models is a singleton), yielding Eq. (12).

4. Conclusion

In this work, we present the complexity formula for atomic partial PI models, the building blocks of non-atomic PI models. We employ the hypercube method for analyzing the complexity of PI models. The following is the discovered equation for the complexity of atomic partial PI models:

$$\omega_{ap} = \prod_{i=1}^{n}(D_i - 1) + \sum_{j=1}^{m}(D_{(B_j)} - 1), \quad \text{(Eq. (19) of Theorem 35)}$$

Note that the term $\prod_{i=1}^{n}(D_i - 1)$ that corresponds to the number of independent joint parameters is the same as that in the full PI model formula (Eq. (17)). This product term expresses the cardinality of the joint space of every variable but its dimension reduced by one. On the other hand, the other term $\sum_{j=1}^{m}(D_{(B_j)} - 1)$ is different from $\sum_{i=1}^{n}(D_i - 1)$ in Eq. (12). Both summation terms represent the number of marginal parameters for either each variable X_i (in the case of Eq. (12)) or each block B_j (in the case of Eq. (19)) that are needed to be specified. The product term $\prod(D_i - 1)$ represent the number of parameters associated with the collective dependency, while the summation terms $\sum(D_i - 1)$ and $\sum(D_{(B_j)} - 1)$ represent the number of parameters related to the marginal independency. It is clear why the product terms are the same for both full and atomic partial PI formula. This is because the collective dependency is the property that is involved only with the collection of *all* variables and, therefore, the collective dependency does not see the types of PI model. On the other hand, the second terms differ since in a full PI model, marginal independency holds for any pair of variables while in an atomic partial PI model, the independency holds for blocks.

Note that this PI model complexity denotes the *upperbound* of the number of independent parameters. This implies that the *minimum* number of parameters *required* for specifying a specific PI model can differ from model to model depending on the domain-specific knowledge available, because the knowledge may impose constraints that are special to that domain and that make the number of required parameters further reduced. For example, if the distribution of a (sub)domain is known to be uniform, then the number of independent parameters of the (sub)domain should be much smaller than the case without this domain knowledge.

The study on the complexity of PI models leads to a new generation of an algorithm equipped with a better scoring metric applicable to learning PI models. This means, from the perspective of using a learned model, the model allows faster inference, more accurate results, and requires less space to represent and to store parameters during inference, as well as provides more expressive power with a compact form. Future studies will focus on revealing the *practical* importance of the theory of learning PI models beyond its theoretical significance as a necessary component of the theory of learning probabilistic networks and its special importance on learning domains of critical areas.

Acknowledgement

The authors are grateful to the anonymous reviewers for their comments on this paper. Financial support from NSERC, Canada is acknowledged.

References

[1] W. Lam, F. Bacchus, Learning Bayesian networks: an approach based on the MDL principle, Computational Intelligence 10 (3) (1994) 269–293.
[2] G. Cooper, E. Herskovits, A Bayesian method for the induction of probabilistic networks from data, Machine Learning 9 (1992) 309–347.
[3] D. Heckerman, D. Geiger, D. Chickering, Learning Bayesian networks: the combination of knowledge and statistical data, Machine Learning 20 (1995) 197–243.
[4] N. Friedman, K. Murphy, S. Russell, Learning the structure of dynamic probabilistic networks, in: G. Cooper, S. Moral (Eds.), Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufman, Madison, WI, 1998, pp. 139–147.

[5] D. Chickering, D. Geiger, D. Heckerman, Learning Bayesian networks: search methods and experimental results, in: Proceedings of the Fifth International Workshop on Artificial Intelligence and Statistics, Society for AI and Statistics, Ft. Lauderdale, FL, USA, 1995, pp. 112–128.
[6] Y. Xiang, S. Wong, N. Cercone, Critical remarks on single link search in learning belief networks, in: E. Horvitz, F. Jensen (Eds.), Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence, Morgan Kaufman, Portland, OR, USA, 1996, pp. 564–571.
[7] Y. Xiang, S. Wong, N. Cercone, A 'microscopic' study of minimum entropy search in learning decomposable Markov networks, Machine Learning 26 (1) (1997) 65–92.
[8] Y. Xiang, J. Hu, N. Cercone, H. Hamilton, Learning pseudo-independent models: analytical and experimental results, in: H. Hamilton (Ed.), Advances in Artificial Intelligence, Springer, Montréal, QC, Canada, 2000, pp. 227–239.
[9] Y. Xiang, Towards understanding of pseudo-independent domains, in: Poster Proceedings of the Tenth International Symposium on Methodologies for Intelligent Systems, Charlotte, NC, USA, 1997.
[10] Y. Xiang, J. Lee, N. Cercone, Towards better scoring metrics for pseudo-independent models, International Journal of Intelligent Systems 19 (8) (2004) 749–768.
[11] Y. Xiang, J. Lee, N. Cercone, Parameterization of pseudo-independent models, in: Proceedings of the Sixteenth Florida Artificial Intelligence Research Society Conference, AAAI Press, St. Augustine, FL, USA, 2003, pp. 521–525.
[12] A. Kolmogorov, Foundations of the Theory of Probability, Chelsea, New York, NY, USA, 1950.
[13] A. Dawid, Conditional independence in statistical theory, Journal of the Royal Statistical Society, Series B 41 (1) (1979) 1–31.
[14] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufman, San Mateo, CA, USA, 1988.
[15] S. Lauritzen, Lecture Notes in Contingency Tables, Institute for Electronic Systems, Department of Mathematics and Computer Science, University of Aalborg, Aalborg, Denmark, 1989.
[16] R. Bouckaert, Bayesian belief networks and conditional independencies, Tech. Rep. RUU-CS-92-36, Utrecht University, Utrecht, The Netherlands, 1992.
[17] M. Studený, Probabilistic Conditional Independence Structures, Springer, 2005.
[18] Y. Xiang, A characterization of single-link search in learning belief networks, in: P.C.H. Motoda, R. Mizoguchi, H. Liu (Eds.), Proceedings of Pacific Rim Knowledge Acquisition Workshop, Singapore, 1998, pp. 218–233.